

RESEARCH ARTICLE

Open Access



Comprehensive genomic analysis of the *CNGC* gene family in *Brassica oleracea*: novel insights into synteny, structures, and transcript profiles

Kaleem U. Kakar^{1,2†}, Zarqa Nawaz^{2,3†}, Khadija Kakar⁴, Essa Ali¹, Abdulwareth A. Almoneafy⁵, Raqeeb Ullah⁶, Xue-liang Ren^{2,8*} and Qing-Yao Shu^{1,7*}

Abstract

Background: The cyclic nucleotide-gated ion channel (CNGC) family affects the uptake of cations, growth, pathogen defence, and thermotolerance in plants. However, the systematic identification, origin and function of this gene family has not been performed in *Brassica oleracea*, an important vegetable crop and genomic model organism.

Results: In present study, we identified 26 *CNGC* genes in *B. oleracea* genome, which are non-randomly localized on eight chromosomes, and classified into four major (I-IV) and two sub-groups (i.e., IV-a and IV-b). The *BoCNGC* family is asymmetrically fractionated into the following three sub-genomes: least fractionated (14 genes), most fractionated-I (10), and most fractionated-II (2). The syntenic map of *BoCNGC* genes exhibited strong relationships with the model *Arabidopsis thaliana* and *B. rapa* *CNGC* genes and provided markers for defining the regions of conserved synteny among the three genomes. Both whole-genome triplication along with segmental and tandem duplications contributed to the expansion of this gene family. We predicted the characteristics of *BoCNGCs* regarding exon-intron organisations, motif compositions and post-translational modifications, which diversified their structures and functions. Using orthologous *Arabidopsis* *CNGCs* as a reference, we found that most *CNGCs* were associated with various protein–protein interaction networks involving *CNGCs* and other signalling and stress related proteins. We revealed that five microRNAs (i.e., *bol-miR5021*, *bol-miR838d*, *bol-miR414b*, *bol-miR4234*, and *bol-miR_new2*) have target sites in nine *BoCNGC* genes. The *BoCNGC* genes were differentially expressed in seven *B. oleracea* tissues including leaf, stem, callus, silique, bud, root and flower. The transcript abundance levels quantified by qRT-PCR assays revealed that *BoCNGC* genes from phylogenetic Groups I and IV were particularly sensitive to cold stress and infections with bacterial pathogen *Xanthomonas campestris* pv. *campestris*, suggesting their importance in abiotic and biotic stress responses.

Conclusion: Our comprehensive genome-wide analysis represents a rich data resource for studying new plant gene families. Our data may also be useful for breeding new *B. oleracea* cultivars with improved productivity, quality, and stress resistance.

Keywords: Abiotic and biotic stress, Ion channels, *CNGC*, Expression pattern, *Brassica oleracea*, Evolution, RNA-seq, qRT-PCR analysis

* Correspondence: renxuel@126.com; qyshu@zju.edu.cn

†Equal contributors

²Molecular Genetics Key Laboratory of China Tobacco, Guizhou Academy of Tobacco Science, Guiyang 550081, China

¹State Key Laboratory of Rice Biology, Institute of Crop Science, Zhejiang University, Hangzhou 310058, China

Full list of author information is available at the end of the article



Background

Calcium is a universal secondary messenger that participates in multiple eukaryotic signalling pathways [1]. In plants, Ca^{2+} signal transduction via calcium-conducting channels is an important mechanism for transducing the signals derived from diverse environmental and developmental stimuli [2, 3]. Additionally, signal transductions contribute to growth, plant biotic interactions, and responses to hormones, light, and salt stress [4]. Cyclic nucleotide-gated ion channels (CNGCs) are components of Ca^{2+} -conducting signal transduction pathways [5]. They are Ca^{2+} -permeable cation-conducting channels that transport sodium, calcium, and potassium cations across membranes. Localized in the plasma membrane [6, 7], vacuole membrane [8], or nuclear envelope [9], CNGCs are controlled from inside the cell by secondary messengers such as Ca^{2+} /calmodulin (CaM) and cyclic nucleotide monophosphates (cNMPs; 3',5'-cAMP and 3',5'-cGMP) [3, 6, 10, 11]. The CNGCs are hypothesized to be involved in the uptake of both essential and toxic cations, Ca^{2+} signalling, development, pollen fertility and tip growth, gravitropism, leaf senescence, innate immunity, pathogen defence, and abiotic stress tolerance [6, 12–15].

The application of bioinformatics tools (for genes/proteins prediction and phylogenetic analysis), and experimental approaches (gene expression, mutant analysis and overexpression in yeast/*Escherichia coli*) have led to the identification, characterization, and functional analysis (in exceptional cases) of CNGC family genes in important plant species, including *Arabidopsis thaliana* [5], rice [16], tomato [17], pear [18], and *Physcomitrella patens* [19]. Researchers have only recently started to investigate the evolution, function (and underlying regulatory mechanism) of plant CNGCs, as well as their phylogenetic relationships with other channels. Briefly, plant CNGCs are characterised by conserved structural components, including a short cytosolic N-terminus, six transmembrane helices (S1–S6) with a pore-forming region between S5 and S6, and a cytosolic C-terminus containing a cNMP-binding domain (CNBD). The CNBD is the most conserved region of CNGCs carrying a plant CNGC-specific motif spanning the phosphate-binding cassette (PBC) and hinge region, which mediates channel gating by cAMP and/or cGMP [3, 20]. A latest study of the *A. thaliana* CNGC12 gene suggested that plant CNGCs have multiple CaM-binding domains (CaMBDs) at cytosolic N- and C-termini [3]. Moreover, channel functionality depends on CaM binding to the conserved isoleucine–glutamine (IQ) motif in the C-terminus of the channel, indicating CaM positively and negatively regulates CNGCs [3]. Studies on individual isoforms and the *A. thaliana* CNGC family revealed that plant CNGC genes may be functionally distinguished in a group-dependent manner. For example, *AtCNGC19* and

AtCNGC20, which belong to Group IV-a, are involved in salt stress responses [8]. Additionally, *AtCNGC2* and *AtCNGC4*, which are Group IV-b members, affect disease resistance against various pathogens and thermotolerance [21, 22]. Mumtaz et al. [4, 17] recently concluded that Group IV-b CNGC genes regulate different types of resistance against diverse pathogens in tomato. It is unclear whether this also applies to other plant species.

Brassica oleracea ($2n = 18$) is a member of the family *Brassicaceae* (approximately 338 genera and 3709 species), which consists of many important vegetable and oilseed crops, including brussels sprout, kohlrabi, and kale [23]. Among the cultivated species, *B. oleracea* exhibits the largest genetic and morphological diversity, making it highly adaptable to different environments. Sexually compatible *B. oleracea* crops, such as cabbage, cauliflower, and broccoli, are valued for their economic, nutritional, and potent anticancer properties [24]. The whole-genome sequence of this plant species was recently published [24], which enabled us to study the *B. oleracea* CNGC family. We used in silico and experimental approaches to identify, characterise, and functionally verify CNGC gene family members. We applied multiple tools and programs to complete in-depth analyses of each CNGC gene family member, including an analysis of the physiological and biochemical properties of the encoded proteins. Our objective was to elucidate the diversification, expansion, and evolution of the CNGC gene family. Furthermore, we investigated CNGC expression patterns to clarify the mechanisms underlying their responses to biotic and abiotic stresses, and to identify novel genes potentially useful for breeding.

Results

Genome-wide identification of CNGC genes in *Brassica oleracea*

For a complete overview of the *B. oleracea* CNGC gene family, we used the 20 *A. thaliana* CNGC genes as queries in BLAST searches of the Ensembl Plants database. Out of the 34 non-redundant putative gene sequences retrieved, eight gene accessions with truncated sequences or lacking CNGC-specific domains (CNBD and transmembrane) were eliminated from analyses (Additional file 1). Finally, 26 CNGC genes containing both essential domains (PF00520/PF07885 and PF00027) and a CNGC-specific motif were identified in the *B. oleracea* genome (i.e., *BoCNGC1–26*). Of the 26 *BoCNGC* genes identified in the latest genome assembly version in Ensembl Plants, 16 and 24 were detected in earlier versions from Bolbase (v.1.3) and GenBank (v.2.1) respectively (Table 1).

The physiological and biochemical properties of the 26 *BoCNGC* proteins were determined by computing different parameters, and are tabulated in Table 1. These proteins varied in length from 558 to 789 amino acids,

Table 1 Properties of 29 BoCNGC genes identified in *Brassica oleracea*

Name	Genes ID (Ensemble v.2.1)	brassicadb (v.1.3)	GeneBank	Transcript (bp)	Protein length (aa)	MolWt. (kDa)	pI	GRAVY	Aliphatic index	II	Total atoms	Ave. residue Wt. (g/mol)	Net charge
BoCNGC1	Bo3g054400	BoI010786	LOC106335861	2115	704	81.33	9.18	-0.212	90.88	47.95	11,473	115.5	22
BoCNGC2	Bo4g009240	BoI000930	LOC106342665	2112	703	81.09	9.78	-0.139	87.92	37.21	11,462	115.3	33.5
BoCNGC3	Bo4g158880	BoI026503	LOC106336663	1950	649	75.08	9.89	-0.023	99.14	39.96	10,654	115.7	39
BoCNGC4	Bo8g100190	BoI045794	LOC106312466	2214	737	84.58	9.74	-0.177	92.51	49.74	11,989	114.8	31
BoCNGC5	Bo4g154120	BoI041958	LOC106342509	2238	745	85.27	9.85	-0.193	89.69	50.69	12,073	114.5	36.5
BoCNGC6	Bo7g114580	BoI033673	LOC106306102	2226	741	84.49	9.53	-0.169	91.61	48.38	11,959	114.0	25.5
BoCNGC7	Bo2g028960	BoI015575	LOC106326694	2166	721	82.52	9.35	-0.122	89.28	50.45	11,653	114.4	26
BoCNGC8	Bo5g021460	BoI038188	LOC106294829	2226	741	85.08	9.17	-0.266	86.22	51.02	11,997	114.8	19.5
BoCNGC9	Bo4g071760	BoI013404	LOC106337043	2055	684	79.43	10.08	-0.11	91.23	49.17	11,253	116.1	38
BoCNGC10	Bo1g013250	BoI001429	LOC106308171	2184	727	84.17	8.8	-0.187	91.21	45.73	11,876	115.8	17
BoCNGC11	Bo4g154790	BoI041907	LOC106340005	2187	728	83.84	9.12	-0.096	93.48	46.34	11,863	115.2	21.5
BoCNGC12	Bo8g099660	BoI044680	LOC106309975	2202	733	84.52	9.29	-0.13	94.69	48.3	11,982	115.3	23
BoCNGC13	Bo9g165600	BoI030411	LOC106316400	2106	701	79.63	8.23	-0.149	84.86	45.08	11,147	113.6	12
BoCNGC14	Bo8g076590	BoI008733	LOC106311631	2121	706	81.58	8.36	-0.247	86.08	52.08	11,403	115.6	15
BoCNGC15	Bo2g042020	BoI037445	LOC106327817	2088	695	80.24	8.25	-0.176	91.12	55.15	11,285	115.5	13.5
BoCNGC16	Bo9g116160	BoI038844	LOC106319027	2097	698	80.78	8.19	-0.192	90.6	55.83	11,357	115.7	12
BoCNGC17	Bo9g164130	BoI06317764	LOC106317764	2151	716	81.93	9.81	0.018	94.61	54.04	11,569	114.4	38.5
BoCNGC18	Bo3g070140	BoI06335359	LOC106335359	2160	674	76.69	8.38	-0.121	89.96	50.37	10,797	114.0	13.5
BoCNGC19	Bo3g070160		1677	558		63.94	7.26	-0.187	87.63	45.22	8953	114.6	5
BoCNGC20	Bo5g122720	BoI06343117	LOC106343117	2262	753	85.98	9.8	-0.075	92.19	52.29	12,121	114.2	32
BoCNGC21	Bo1g119310	BoI06344554	LOC106344554	2280	759	86.21	9.81	-0.067	90.59	47.96	12,147	113.6	32.5
BoCNGC22	Bo1g119340		2205	734		82.99	9.2	0.033	93.37	47.16	11,692	113.1	28
BoCNGC23	Bo1g079060		2370	789		89.00	10.18	-0.055	95.64	46.62	12,604	112.8	42
BoCNGC24	Bo1g119320		2232	743		85.06	10.06	-0.103	90.93	53.01	11,990	114.5	37
BoCNGC25	Bo5g122750	BoI06293228	LOC106293228	2346	781	89.78	10.11	-0.176	90.73	47.88	12,686	115.0	34.5
BoCNGC26	Bo5g122740	BoI06293974	LOC106293974	2235	744	85.72	9.3	-0.231	85.44	53.09	12,036	115.2	21.5

with an average of 717 amino acids. The ProtParam tool revealed that there was a considerable range in BoCNGC residue weight (112.795–116.128 g/mol) and molecular weight (63.938–89.775 kDa) depending on the number of atoms present. The computed average *pI* of majority of BoCNGC proteins was relatively high (range 8.23 to 10.18), signifying that these proteins are localized to membranes, and will supposedly participate in basic buffers. The BoCNGC19, which had *pI* than 7.4, indicate that this protein likely participate in the acidic buffers. Approximately one third of BoCNGC proteins had a low net charge (<17), while other are composed of more charged amino acids. Nearly all BoCNGC were hydrophilic, with BoCNGC17 and BoCNGC22 being slightly hydrophobic, which endorses its multifaceted role in cellular membrane transport. According to the instability index (II), only two proteins were stable in test tubes, namely BoCNGC2 and BoCNGC3. Aliphatic index showed that most BoCNGC proteins were thermostable at a wide temperature ranges, similar to other globular proteins.

Phylogenetic analysis of BoCNGC genes

Multiple sequence alignments and a maximum likelihood phylogenetic tree constructed between BoCNGCs and AtCNGCs were used to determine the similarity and homology between the *B. oleracea* and *A. thaliana* CNGC families. To strengthen the phylogenetic analysis, we identified and included 29 CNGC homolog genes from sister specie *Brassica rapa* (*BrCNGCs*) in current analysis. The sequence alignment revealed high similarity between the amino acid sequences of the three species, especially at the conserved domain regions (Additional file 2). The topology of the inferred maximum likelihood scoring tree revealed that the *BoCNGC* gene family can be divided into four major groups (i.e., Groups I–IV), which are based on the *A. thaliana* groups (Fig. 1) [5]. Groups I–III are monophyletic, while Group IV is sub-divided into two distinct clades (i.e., Groups IV-a and IV-b). Group IV contains 12 *BoCNGC* genes, while the other groups contain three to six members. Moreover, individual phylogenetic trees that were constructed based on the aligned *B. oleracea* and *A. thaliana* CNGC proteins produced similar clustering patterns (Additional files 3 and 4).

Chromosomal distribution and diversification of BoCNGC genes

The 26 *BoCNGC* genes were mapped onto *B. oleracea* chromosomes, and the position of each locus was determined. These genes were randomly distributed across the genome, and were detected on eight of nine chromosomes (i.e., C1–5 and C7–C9). The *BoCNGC* genes were unevenly distributed, with some chromosomes (i.e., C1

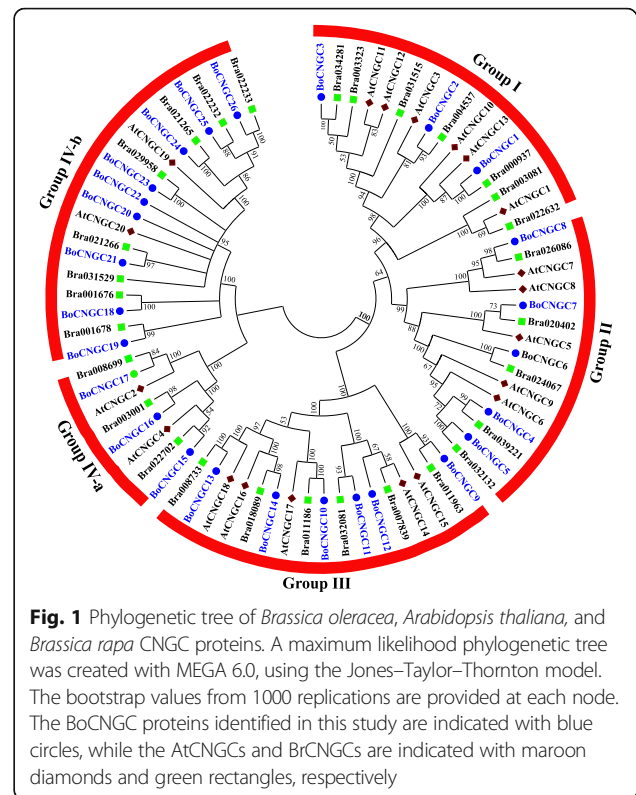


Fig. 1 Phylogenetic tree of *Brassica oleracea*, *Arabidopsis thaliana*, and *Brassica rapa* CNGC proteins. A maximum likelihood phylogenetic tree was created with MEGA 6.0, using the Jones–Taylor–Thornton model. The bootstrap values from 1000 replications are provided at each node. The *BoCNGC* proteins identified in this study are indicated with blue circles, while the *AtCNGCs* and *BrCNGCs* are indicated with maroon diamonds and green rectangles, respectively

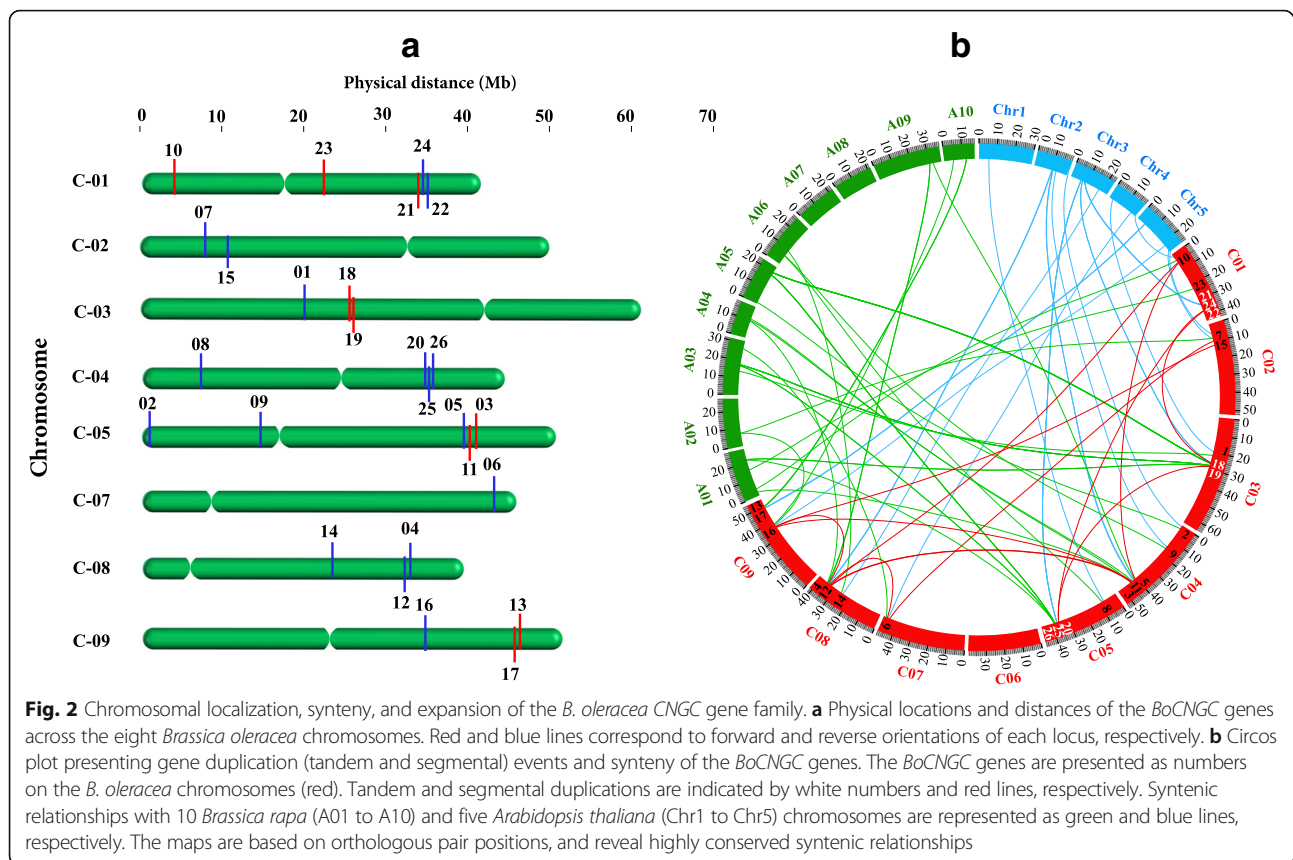
and C5) carrying five genes, while the rest had fewer genes (e.g. C7). Chromosome 6 did not carry any of the *BoCNGC* genes (Fig. 2a).

Gene duplication events

Gene family expansion occurs via the following three mechanisms: tandem duplication, segmental duplication, and whole-genome duplication [25]. We investigated gene duplication events to clarify the genome expansion mechanism of the *B. oleracea* *BoCNGC* superfamily. An evaluation of the physical distance between *BoCNGC* gene loci revealed that eight genes (i.e., *BoCNGC18/BoCNGC19*, *BoCNGC21/BoCNGC22/BoCNGC24*, and *BoCNGC20/BoCNGC25/BoCNGC26*) were tandemly duplicated. These genes were detected on C3, C1, and C5, respectively. The data obtained from the Plant Genome Duplication Database revealed that 13 *BoCNGC* genes distributed across the *B. oleracea* genome were associated with segmental duplications (Fig. 2b). The *BoCNGC* gene clusters likely formed via tandem and segmental duplication events may have expanded and enhanced the functional diversity of the gene family.

Comparative syntenic and evolutionary analyses of orthologous CNGC gene pairs

The *B. oleracea* and *B. rapa* genomes are currently divided into three sub-genomes, namely LF (least fractionated), MF-I (most fractionated), and MF-II [26]. We



observed that the *B. oleracea* LF sub-genome contains the most *BoCNGC* genes (14), followed by sub-genomes MF-I (10) and MF-II (2) (Additional file 5). Because of a *Brassica*-lineage specific whole-genome triplication (WGT) [27], each *A. thaliana* *CNGC* gene was expected to generate three *Brassica* copies. However, there were 20 *A. thaliana* *CNGC* genes, 26 *B. oleracea* *CNGC* genes, and 29 *B. rapa* *CNGC* genes. To detect the retention or loss of *CNGC* genes after a WGT event, the syntenic map of *BoCNGC* genes with the model *A. thaliana* and *B. rapa* *CNGC* genes provided markers for defining the regions of conserved synteny among the three genomes (Fig. 2b). Compared with the ancestral *Brassicaceae* blocks (A to X) in *A. thaliana*, the synteny of 15 *AtCNGC* genes was preserved in *Brassica* species, based on the number of corresponding genes. Ten of the 20 *AtCNGC* genes were retained as a single copy in the equivalent blocks of both *Brassica* species. Three *AtCNGC* genes (i.e., AT2G23980, AT2G24610, and AT5G54250) located on the I and W syntenic blocks, were preserved as two copies in *Brassica* genomes, which were asymmetrically fractionated into three sub-genomes. Two *AtCNGC* genes (i.e., AT3G17690 and AT3G17700) in the F syntenic block were retained as three copies in each species. Two extra gene copies (i.e., *BoCNGC20* and *BoCNGC22*) were located on potential

overlap/tandem repeat regions of the *B. oleracea* genome, thus producing phylogenetic cluster IV-b. Approximately 25 *B. oleracea* *CNGC* genes and 24 *B. rapa* *CNGC* genes exhibited clear syntenic relationships among the three species. Two gene pairs (i.e., *BoCNGC3* and *BoCNGC23*; Bra034281 and Bra029958) were not part of an *A. thaliana* syntenic block (Additional file 6), suggesting that these genes originated after the divergence from *A. thaliana*. The remaining four *B. rapa* genes were likely generated after the speciation event. In addition, 11 *BoCNGC* genes exhibited strong syntenic relationships with the genes from other plant species, implying this gene family is important for plant growth, development, and stress resistance (Additional file 6).

The orthologous *CNGC* gene pairs between the *B. oleracea* and *A. thaliana* genomes were used to estimate the K_a , K_s , and K_a/K_s values (Table 2). The mean K_a/K_s value of all orthologous gene pairs in the *B. oleracea* *CNGC* gene family was 1.98. Most of the *BoCNGC* genes had K_a/K_s ratios greater than 1. Additionally, the minimum and maximum K_a/K_s ratios were 1.05 (*BoCNGC26*) and 7.7 (*BoCNGC6*), respectively. These findings indicate that the *BoCNGC* gene family is under positive selection pressure, and might preferentially conserve functions and structures under this selective pressure.

Table 2 Comparative analysis of Ka, Ks and Ka/Ks values for *CNGC* gene pairs between *B. oleracea* compared to *A. thaliana*. Ka/Ks ratio greater than 1 indicates positive selection, a ratio less than 1 indicates functional constraint, and a Ka/Ks ratio equal to 1 indicates neutral selection

<i>A. thaliana</i> genes	<i>B. oleracea</i> genes	KA	KS	KA/KS
AtCNGC13	BoCNGC1	0.137	0.032	4.303387233
AtCNGC3	BoCNGC2	0.147	0.029	5.117236906
AtCNGC6	BoCNGC4	0.136	0.052	2.615208996
	BoCNGC5	0.167	0.048	3.468082016
AtCNGC9	BoCNGC6	0.188	0.024	7.768521972
AtCNGC5	BoCNGC7	0.124	0.040	3.098687155
AtCNGC7	BoCNGC8	0.111	0.033	3.40004813
AtCNGC15	BoCNGC9	0.147	0.060	2.456018066
AtCNGC17	BoCNGC10	0.113	0.034	3.357834045
AtCNGC14	BoCNGC11	0.133	0.031	4.245278743
	BoCNGC12	0.145	0.040	3.651430365
AtCNGC18	BoCNGC13	0.094	0.048	1.949453718
AtCNGC16	BoCNGC14	0.111	0.066	1.676233706
AtCNGC4	BoCNGC15	0.101	0.025	4.039237878
	BoCNGC16	0.103	0.034	3.056103924
AtCNGC2	BoCNGC17	0.118	0.029	4.091069466
AtCNGC19	BoCNGC18	0.246	0.126	1.950211367
AtCNGC20	BoCNGC19	0.202	0.178	1.133449904
	BoCNGC20	0.136	0.049	2.79662626
	BoCNGC21	0.099	0.041	2.44174101
AtCNGC19	BoCNGC22	0.202	0.119931313	1.680530313
	BoCNGC24	0.146	0.081	1.792823624
	BoCNGC25	0.146	0.081	1.792823624
AtCNGC20	BoCNGC26	0.131	0.125	1.054276842

Domain architecture and alignment of BoCNGC proteins

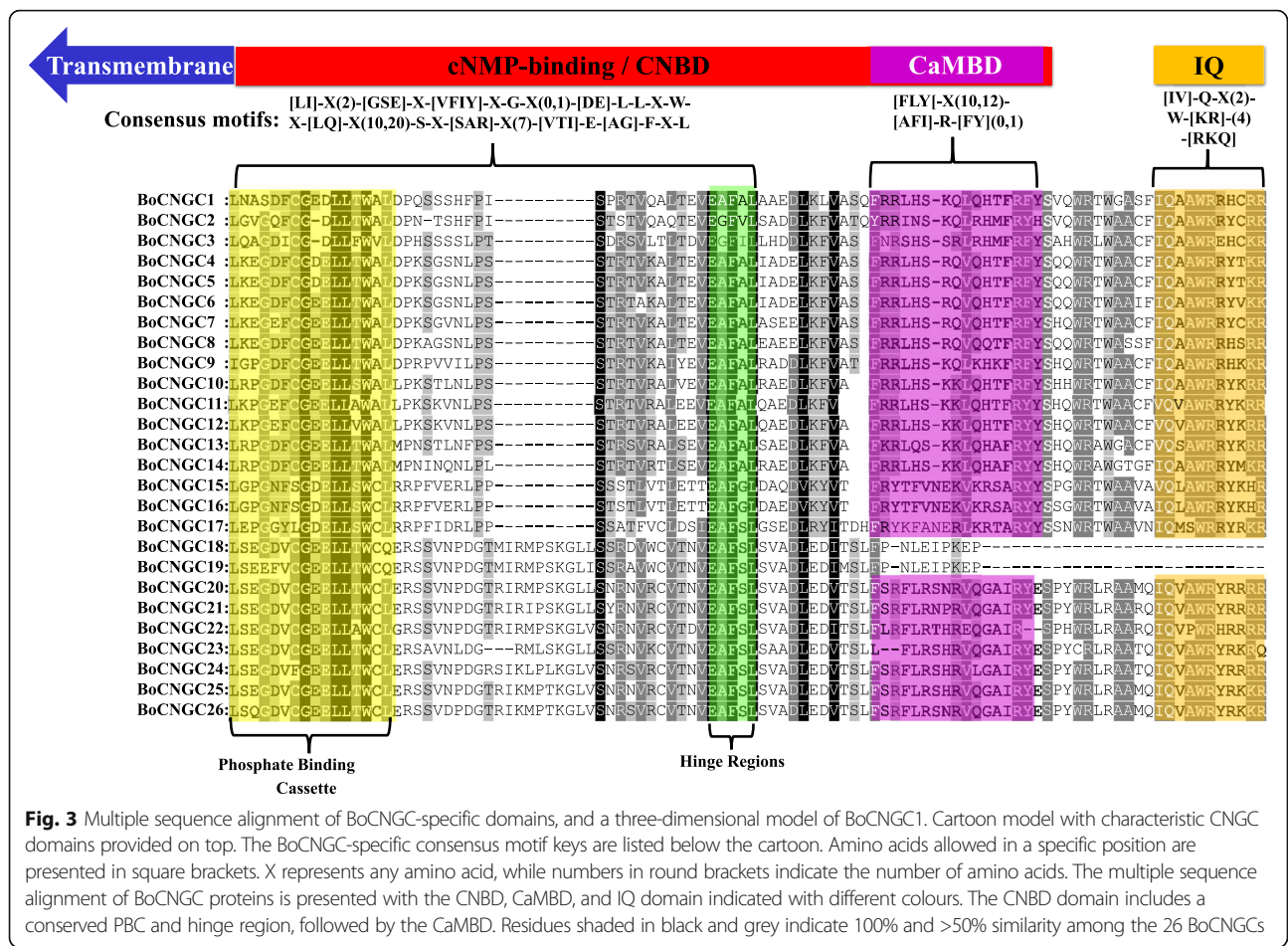
Domain composition analyses revealed that BoCNGC proteins contain two primary domains, namely CNBD and TM (Additional file 7). The sequence alignment of 26 BoCNGCs indicated that the two most conserved regions within the CNBD domain are a PBC, and an adjacent hinge region (Fig. 3; Additional file 8). The following highly conserved consensus motif was identified: [LI]-X(2)-[GSE]-X-[VFIY]-X-G-X(0,1)-[DE]-L-L-X-W-X-[LQ]-X(10,20)-S-X-[SAR]-X(7)-[VTI]-E-[AG]-F-X-L. This sequence can be used to classify newly annotated or un-annotated candidate sequences as *Brassica* CNGCs. Additionally, there was a relatively conserved IQ domain and a less conserved CaMBD adjacent to a CNBD present in 24 of the 26 BoCNGC proteins. Two proteins (i.e., BoCNGC18 and BoCNGC19) were observed to lack the CaMBD and IQ domains because their sequences are truncated at the C-terminal end of the CNBD. A high

sequence divergence was noted among different groups, particularly between members of Sub-groups IV-a and IV-b. For example, the CaMBD [FLY[-X(10,12)-[AFI]-R-[FY](0,1), was not particularly conserved between Group IV-b and the other groups. However, the IQ motif [IV]-Q-X-X-W-R-X-X-X-[RKQ] was relatively conserved among the BoCNGC proteins (Fig. 3). Alignments between BoCNGCs, AtCNGCs, and BrCNGCs revealed a high sequence divergence at the C-terminal of the CNBD, in which several Group IV-b members lack the CaMBD and IQ motif (Additional files 9 and 10). Overall, our in silico analyses suggest that ion transport and CNBDs along with the PBC and hinge region are conserved in all three species, and are characteristic of plant CNGCs.

Gene structure and motif composition analysis

To characterise the structural diversity of the *BoCNGC* family members, we analysed the exon–intron organization of individual *BoCNGC* genes. The majority of the *BoCNGC* genes from phylogenetic Groups I–III contained six or seven exons, while the Group IV members had 8–11 exons (Fig. 4). Closely clustered *BoCNGC* genes in the same clades were similar regarding the number of exons and intron lengths. Most of the introns in *BoCNGC* genes were phase 0 introns, which occur in between complete codons. Fifty-four phase 2 introns (i.e., located between the second and third nucleotides of a codon) were observed in the *BoCNGC* family, in which the genes carried two phase 2 introns. The exceptions were *BoCNGC1* and *BoCNGC2*, which contained three phase 2 introns. Only the members of phylogenetic Group IV-b had single phase 1 introns at the terminal end of their sequences. A comparison between the exon–intron organizations of *BoCNGC* genes and the *AtCNGC* genes clustered in the same phylogenetic groups revealed several differences (Additional file 11). Most of the phase 1 introns were present in *AtCNGC* genes, implying that intron loss during evolution resulted in a decrease in the number of introns in *BoCNGC* genes, particularly those in Groups I–III and IV-a.

The BoCNGC protein sequences were used for domain or motif structure analyses with the Multiple Expectation Maximization for Motif Elicitation suite [28]. Ten conserved motifs were identified. According to Pfam codes [29] and WebLogo, only seven motifs (i.e., 1–5, 7, and 10) encode domains with known functions (Fig. 4; Additional files 12 and 13). Motif 2 was the biggest motif encoding a conserved domain, which is probably associated with peptidase_C50, putative aminopeptidase, or DNA polymerase III subunit tau_4. Motifs 1 and 5, which encode a CNBD and an ion transport domain, respectively, were conserved among all BoCNGC family members. The ion transport domain had the most motifs, including motifs 4,



5, 7, and 10. The IQ CaM-binding motif (PF00612) was conserved among BoCNGC family members, with the exception of BoCNGC18, 19, and 22. Group IV proteins contained the fewest functionally annotated motifs, suggesting that the closely related proteins in each group have similar motifs and are also probably functionally similar. The functions of the remaining motifs (i.e., 6, 8, and 9) remain to be determined.

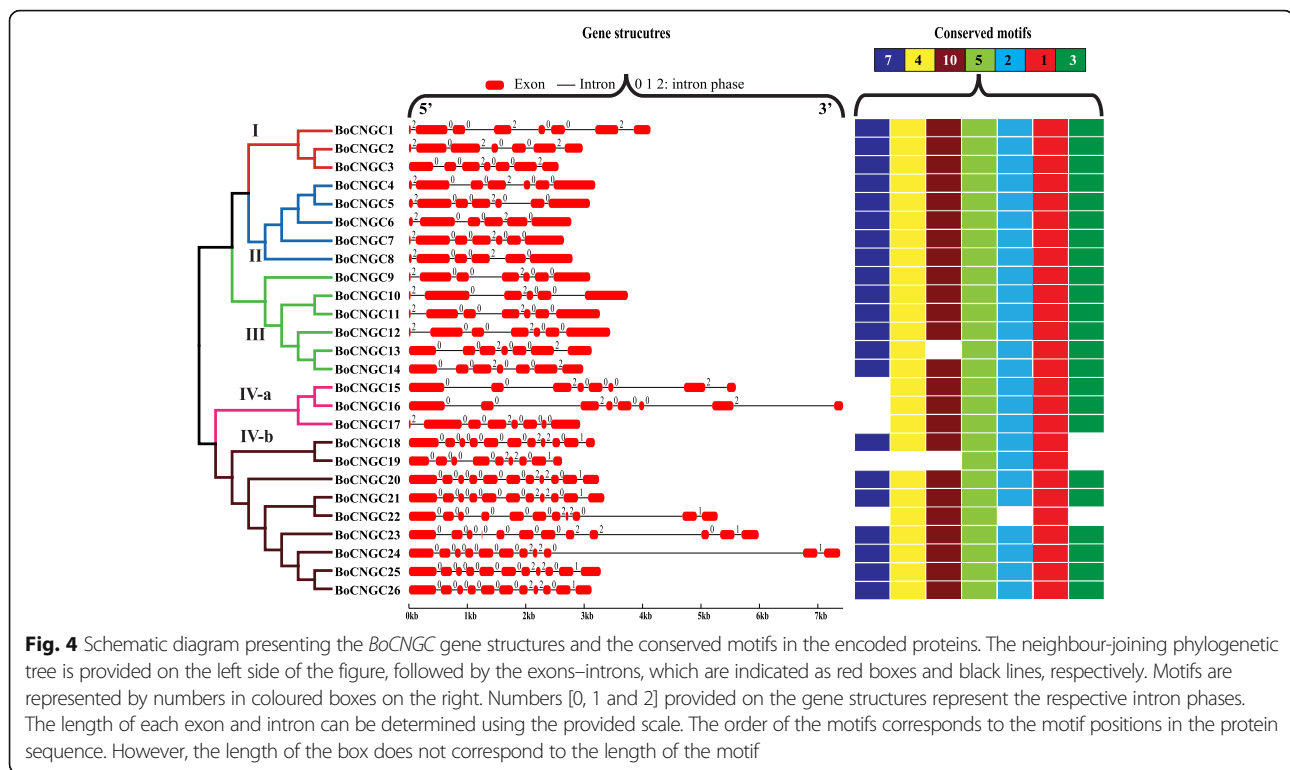
Post-translational modification and phosphorylation of BoCNGC proteins

When BoCNGC protein sequences were analysed using ScanProsite [30], multiple putative phosphorylation sites were revealed. These sites may act as substrates for several kinases, including casein kinase II, protein kinase C, tyrosine kinase, and cAMP/cGMP kinases. Protein kinase C, a family of ten isoenzymes that play a vital role in cellular signal transduction [31], were the most abundant, with 16 sites in BoCNGC4, BoCNGC5, BoCNGC8, and BoCNGC12. Casein kinase II sites, which were the most abundant in Group IV members, are reported to influence different developmental and stress responsive pathways in Arabidopsis [32]. All BoCNGC proteins had multiple

N-myristoylation/N-glycosylation motif sites, which are highly conserved compared with the other PTMs. The lipid modification by N-myristoylation might controls the redox disproportions originating from different stresses in plants [33], while glycosylation is crucial for correct growth [34]. The BoCNGC5 and BoCNGC18 proteins contained the most N-myristoylation (11) and N-glycosylation (10) sites, respectively. Other PTM sites, such as those for amidations, tyrosine kinase, serine- and glutamic acid- rich regions, cell attachment sequences, and leucine zipper patterns, were less conserved and randomly distributed (Table 3). Such phosphorylations deliver effective means to regulate most physiological activities, including metabolism, transcription, DNA replication and repair, cell proliferation [35].

Prediction of functional association network of BoCNGC proteins

To explore the relationships among different BoCNGC proteins, a hypothetical protein–protein interaction network was in silico predicted with the STRING program (accessed in April 2016) [36] and AtPID (*Arabidopsis thaliana* Protein Interactome Database), using using



orthologous *AtCNGCs* as query. The STRING interaction network for the first shell of interactors of *AtCNGC* proteins, supported by confidence score, is presented in Fig. 5a. Fourteen *AtCNGCs*, having 24 orthologs in *B. oleracea*, interact with flagellin-sensitive 2 (i.e., *FLS2* or *MPL12.8*), represented by association in curated databases (confidence score: 0.8). This association was traced to manually curated plant–pathogen interaction pathway imported from the Kyoto Encyclopedia of Genes and Genomes database (Additional file 14). Supported by principal component analysis, a positive interaction (confidence score: 0.154) was observed between *BoCNGC10* and *BoCNGC13*, which are the orthologues of *AtCNGC17* and *AtCNGC18*, respectively. In another interaction network, *BoCNGC1* interacts with *BoCNGC2* and *BoCNGC18–26*, which are orthologues of *AtCNGC13*, 2, 19 and 20 respectively. This interaction is based on protein homology, association in curated human pathways (<http://www.reactome.org/>), or genes encoding these proteins have correlated expression levels. We also observed that the Group IV proteins are associated with constitutive photomorphogenic 1 and CaM proteins (i.e., *CaM4*, *CaM6*, and *CaM7*) (Fig. 5a).

Using orthologous *Arabidopsis* *CNGCs* as query in the *AtPID* uncover more potential interactions between *CNGCs*, and to other proteins, which are validated by experimental data from different assays (Fig. 5b; Additional file 15). The results exhibited strong

interactions of co-expression and gene fusion between *CNGC* functional partners belonging to similar clades. For example, *AtCNGC10* interacted with *AtCNGC1*, 3 and 13, while *AtCNGC17* interacted with *AtCNGC18* as mentioned earlier. *AtCNGC10* interacted with more *CNGCs* than other proteins. In addition, some *CNGCs* (*AtCNGC1*, 5, 6, 9, 10, 13, 17, 18 and 19) interacted with many important signaling and stress related regulatory proteins, including calmodulins. These interactions are supported by data from yeast two-hybrid, and Affinity Capture-MS assays. Five *CNGC* genes (*AtCNGC 1–4*, and 11) were found to have available phenotypes of mutant data from seedlings, leaves and embryos, showing that these genes play important roles in hypersensitivity, pathogen and abiotic stress resistance (Additional file 15).

Additional evidence from experimental/biochemical data detected by protein kinase (MI:0424) and anti tag coimmunoprecipitation (MI:0007) assays in human putative homologs (i.e., Potassium voltage-gated channel 2 and Leucine rich repeat containing 47/Per-Arnt-Sim domain kinase) suggest a functional link between *CNGCs* and *FLS2* [37, 38]. The experimental details and LC-MS/MS, yeast two-hybrid and phosphorylation of peptide arrays of human interacting *KCNH2* and *LRR47/PASK* proteins can be found in supplementary material of Behrends et al. [38]. Using Mating-Based Split Ubiquitin Assays in *A. thaliana*, Chen et al. [39] reported strong, positive (in both 500 μ M methionine

Table 3 The number of predicted post-translational modification sites in BoCNGC protein sequences

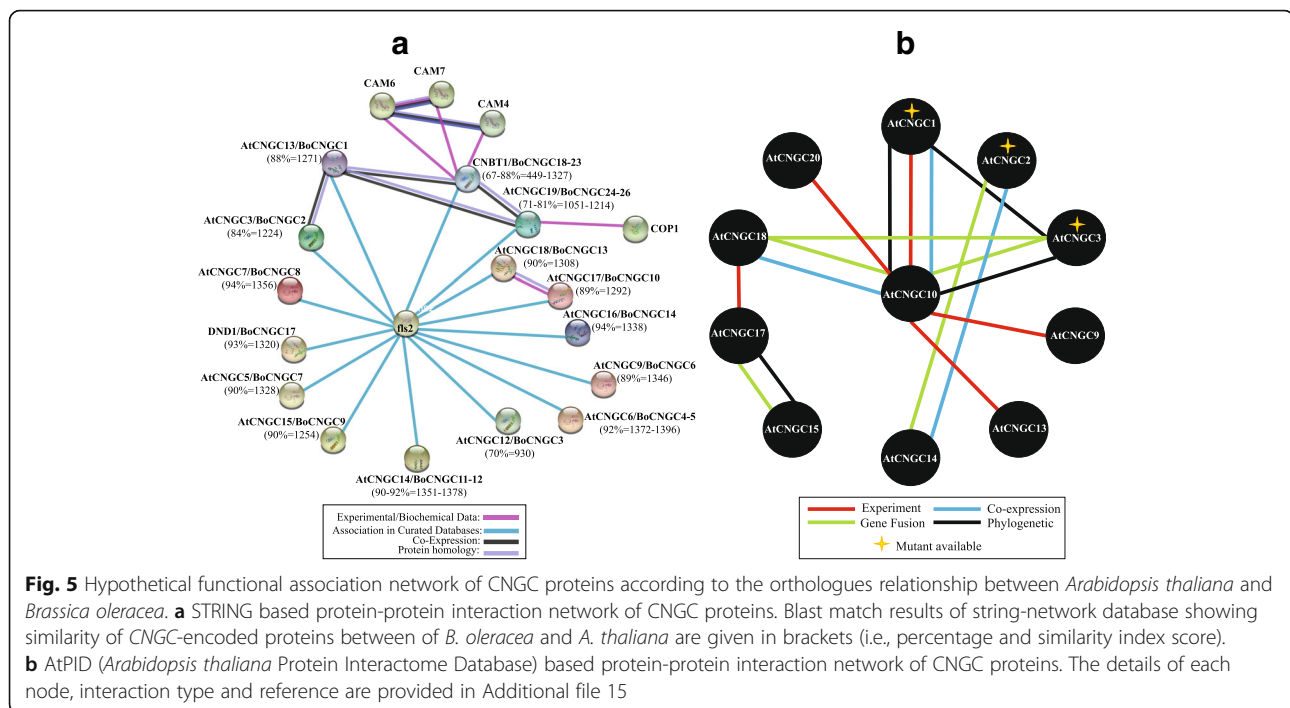
Protein ID	CAMP	CK2	AMD	PKC	ASN	TYR	MYR	RGD	LEU	SER	GLU	ATP
BoCNGC1		7	2	8	7	1	4					
BoCNGC2	2	6		7	4	1	7					
BoCNGC3	2	7		10	4	1	3					
BoCNGC4	4	3		16	5	1	8					
BoCNGC5	3	4		16	4	1	10					
BoCNGC6	2	8		14	6	1	11					
BoCNGC7	1	5	1	9	5	2	7		1			
BoCNGC8	3	6		16	3	1	7					1
BoCNGC9	1	4		12	4	2	8					
BoCNGC10	1	6		12	4	2	7					
BoCNGC11	1	7		15	2	1	5		3			
BoCNGC12	2	9		16	2	1	5					
BoCNGC13	2	8	1	11	7		9		3			
BoCNGC14	1	8		13	5	1	8					
BoCNGC15	1	12	1	8	4		9		1		1	
BoCNGC16	2	13		10	5		8				1	
BoCNGC17	2	9		6	3	1	6					
BoCNGC18	2	11	1	9	10	1	5		1			
BoCNGC19	1	10		9	7	1	3	1	1			
BoCNGC20		13	1	9	2	1	5					
BoCNGC21		12	1	8	2		7	1				
BoCNGC22		11		6	4		5			1		
BoCNGC23	2	11	1	14	3		7					
BoCNGC24		14		15	6		8					
BoCNGC25		13		13	5		6					
BoCNGC26	1	14		6	7		4					

cAMP/cGMP cAMP/cGMP-binding motif profile, *SER* serine-rich region profile, *GLU* glutamic acid-rich region profile, *CAMP* cAMP- and cGMP-dependent protein kinase phosphorylation site; *CK2* casein kinase II phosphorylation site, *AMD* amidation site, *PKC* protein kinase C phosphorylation site, *ASN* N-glycosylation site, *TYR* tyrosine kinase phosphorylation site, *MYR* N-myristoylation site, *RGD* cell attachment sequence, *LEU* leucine zipper pattern, *ATP* ATP/GTP-binding site motif A (P-loop). Numbers given in each cell refer the total count of PTM sites found in each protein

and at least one 150 μ M methionine conditions), and statistically significant interaction between these protein pairs, which are required for polarized tip growth of pollen tube [40]. In another interaction network, BoCNGC1 interacts with BoCNGC2 and BoCNGC18–26, which are orthologues of AtCNGC13, 2, 19 and 20 respectively. Additionally, we observed a weak interaction (confidence score: 0.151) between AtCNGC13 (i.e., orthologues of BoCNGC1) and BRI-associated receptor kinase 1 (BAK1), which was previously observed between AtCNGC17 and BAK1 [41]. Though, it is reported that evidence transfer from one model organism to the other seems feasible approach to study interaction conservation, and it has been implemented in several frameworks already [42]. However, these experimental proofs are essential to support this analysis in *B. oleracea*.

Identification of microRNA target sites

Identifying the targets of the predicted microRNAs (miRNAs) may provide insights into the biological functions of miRNAs influencing plant development, signal transduction, and stress adaptations [43]. We searched for potential miRNA targets in a set of identified *BoCNGC* transcripts using the plant small-RNA target analysis server (psRNATarget) [44]. Using a cut-off threshold of 5 for the search parameters, we identified 14 miRNAs with target sites in 17 *BoCNGC* transcripts, with expectation scores of 1.5–5 (Additional file 16). To decrease the number of false positive predictions, small-RNA/target site pairs with an expectation score and cut-off threshold of 3 were considered. Consequently, five miRNAs with target sites in nine *BoCNGC* genes were identified (Table 4). These miRNAs were localized to the 3' arm of the stem-loop hairpin structure. Unlike bol-miR838d, which has



five target genes, the remaining miRNAs have only one target gene. Moreover, only bol-miR838d has multiple target sites (i.e., complementary regions) on *BoCNGC15* and *BoCNGC16* transcripts. The accessibility of the target site varied from 2.883 (bol-miR838d) to 16.4 (bol-miR5021), where lower values correspond to a greater possibility of contact between the miRNA and target site. Four miRNAs were determined to be involved in cleaving the target transcript, while two miRNAs were predicted to inhibit the translation of target genes.

Gene ontology enrichment analysis

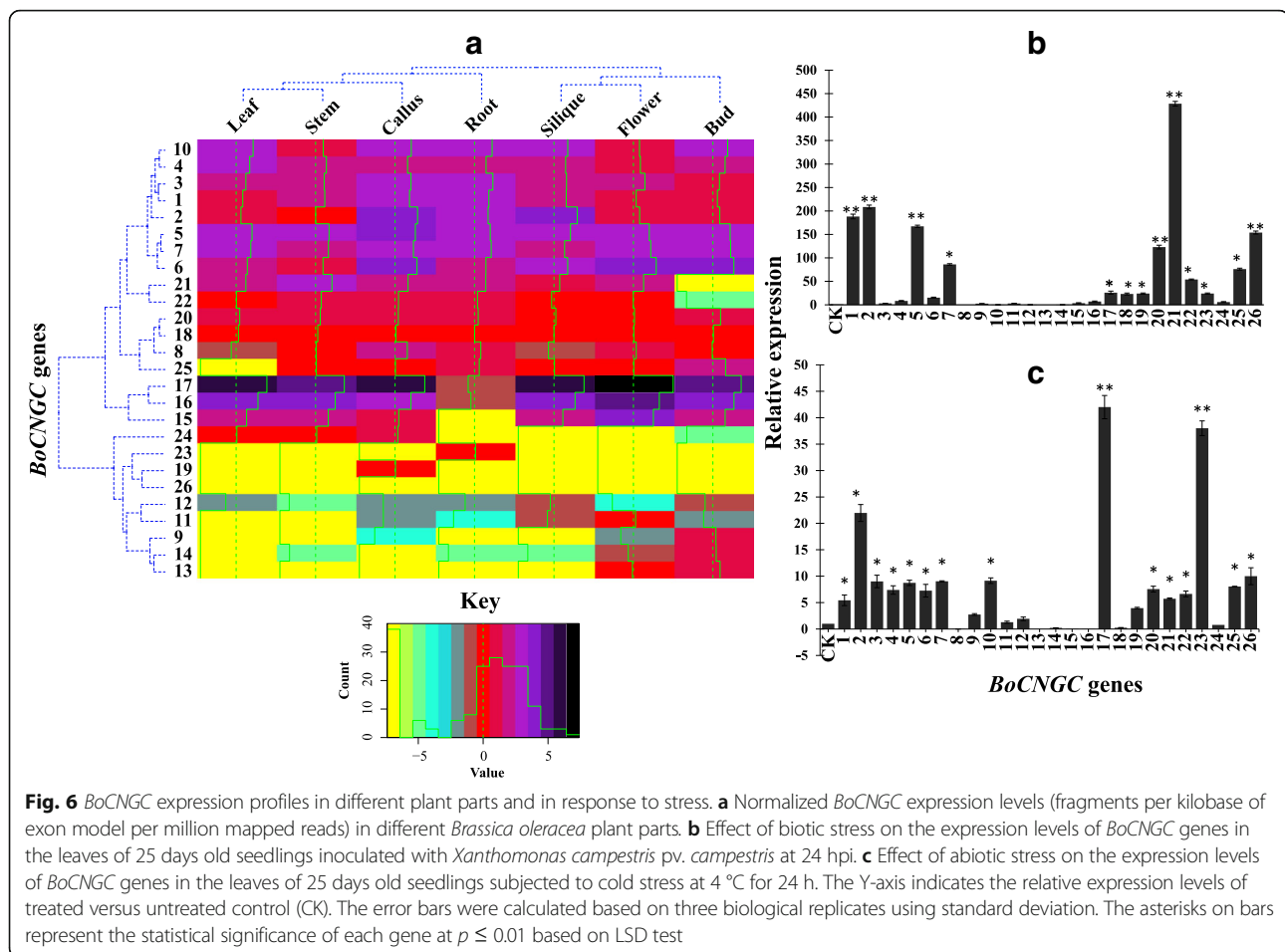
Using Blast2GO (v.3.3.5), we assigned 31 gene ontology (GO) classes to 26 *BoCNGC* genes with BLAST matches to known proteins in the InterPro database. The majority of the genes were assigned to biological process (22), followed by molecular function (7) and cellular components (3). All genes encoded integral membrane components associated with ion channel activity for transmembrane transport. Notably, *BoCNGC1* was associated with salicylic acid biosynthesis, negative regulation of defence responses, regulation of plant-type hypersensitive responses, and responses to chitin. Additionally, *BoCNGC6* was associated with DNA-mediated transformation (Additional file 17).

The level 2 GO enrichment analysis revealed that all 26 *BoCNGC* proteins are integral cell membrane components, with four proteins (i.e., *BoCNGC1*, *BoCNGC4*, *BoCNGC5*, and *BoCNGC17*) forming cell parts, and two proteins (i.e., *BoCNGC4* and *BoCNGC5*) forming

macromolecular complexes (Additional files 18-a and 19). These proteins are involved in cellular processes associated with transport, binding, and transduction (Additional files 18-b and 19). The biological process category at GO level 2 indicated that *BoCNGC1* and *BoCNGC17* are associated with cell death and immune responses to stimuli, while another eight CNGCs, including *BoCNGC19*, are associated with localization (Additional files 18-c and 19). Moreover, we mapped the 26 annotated sequences to reference pathways in the Kyoto Encyclopedia of Genes and Genomes database [45]. Twenty-four of these genes were defined as “cyclic nucleotide gated channels”, and assigned to the “plant-pathogen interaction” pathway (Additional files 14 and 20).

Expression patterns in different plant parts

We investigated the steady-state *B. oleracea* *BoCNGC* expression patterns in seven tissues (i.e., leaf, stem, callus, root, silique, flower, and bud) using Illumina RNA-sequencing data from the Gene Expression Omnibus database. Of the 26 *BoCNGCs*, 19 were expressed at relatively high levels (fragments per kilobase of exon model per million mapped reads value >1) in at least one tissue, including 15 in the roots and siliques, 16 in leaves, and 17 in stems, buds, and flowers. The 19 genes were also expressed in calli (Fig. 6a). Some of the syntenic duplicates have diverged in expression patterns indicating sunfunctionalization. For example, *BoCNGC26* and *BoCNGC19* have very similar expression patterns. But their



expression levels, including the up-regulation of the expression of 10 *BoCNGC* genes in infiltrated seedlings, with the highest levels observed for *BoCNGC21*. This was followed by *BoCNGC2* and *BoCNGC1* from Group I, *BoCNGC5* and *BoCNGC7* from Group II, and *BoCNGC26* and *BoCNGC20* from Group IV-b. Interestingly, none of the Group III and IV-a genes were affected.

We also examined the *BoCNGC* expression levels under cold conditions. The expression of 13 of the 26 *BoCNGC* genes was up-regulated in cold-stressed plants, although the expression levels were lower than the levels induced by Xcc (i.e., biotic stress) (Fig. 6c). The expression levels of genes from Groups I, II, and IV were significantly induced by cold stress, with the highest levels observed for *BoCNGC17* and *BoCNGC23*. In contrast, the Group III *BoCNGCs* were expressed at low levels or not at all under cold conditions. Moreover, most of the duplicated gene pairs and genes encoding interacting proteins produced similar expression patterns (especially in response to Xcc). The exception was *BoCNGC24* whose expression was not significantly up-regulated like its duplicates (i.e., *BoCNGC21* and *BoCNGC22*).

The expression patterns of many *BoCNGCs* under pathogen stress were consistent with the expression patterns of their *Arabidopsis* orthologs obtained from the AtGenExpress project (Additional file 22) [46]. The involvement of group-IV CNGCs in disease resistance and hyper-sensitivity has been documented earlier [21, 22]. However, the cumulative profiles of group-I and IV CNGCs in *Arabidopsis* seedlings showed opposite trend of down-regulation by cold stress at 4 °C for 24 h, showing specie-specific divergence of expression pattern.

Discussion

The *CNGC* gene family has been reported for many agriculturally important plants [17, 18, 20]. However, a genome-wide identification and annotation of *CNGC* genes has not been reported for *B. oleracea*. In this study, we identified 26 *B. oleracea* *CNGC* genes, and determined that the *BoCNGC* gene family is larger than the *CNGC* families of most of the reported crops [4]. The isoelectric point (pI) and charge of a protein is important for solubility, subcellular localization, and interaction, depending on both insertion and deletions

between orthologs, and the ecology of the organism [48]. It is reported that proteins in cytoplasm possess an acidic pI ($pI < 7.4$), nuclear proteins have more neutral pI ($7.4 < pI < 8.1$), while those in membrane have more basic pI [48], where basic residues located on either side of membrane spanning region play a role in the stabilization of the protein in membrane [49]. The net charge of a protein is a fundamental physical property, and its value directly influences the solubility, aggregation, and crystallization of the protein [50]. The 26 BoCNGCs were localized to membranes, greatly varied in physicochemical properties, and will theoretically participate in basic buffers. These variations reflect the changes in protein composition, and their effects on association of receptors with charged ligands, folding and stability, solubilization and precipitation, and selective transport of ions in protein channels [50].

Homologous genes within the same taxonomic group are assumed to exhibit similar structural, functional, and evolutionary properties, which may help clarify the role(s) of *B. oleracea* CNGC genes. Because of the close relationship between *B. oleracea* and *A. thaliana*, the BoCNGC genes were highly similar (>90%) to the corresponding AtCNGC genes regarding plant CNGC-specific domains, amino acid compositions, gene structures, and phylogenetic classifications. Interestingly, we revealed the absence of the CaMBD and IQ domain in BoCNGC18 and BoCNGC19, which raised the possibility that these were abnormal CNGC proteins. However, we found that many of their homologs in *A. thaliana*, pear and *B. rapa* reportedly lack the CaMBD [18]. Similar to other CNGCs, these proteins have regular 3D structural and membrane topologies, with conserved binding sites for cGMP/cAMP. Furthermore, the presence of conserved nickel- and zinc-binding sites suggests that BoCNGC18 and BoCNGC19 may have lost their secondary domains during evolution, but gained functional diversity. Additional research is required to clarify this point.

Proteins undergo post-translational modifications (PTMs), which increase the range of their functions through different mechanisms [51]. The associated PTMs likely affected protein function, localization, and stability, as well as the dynamic interactions with other molecules [52]. Following gene annotations and phylogenetic analyses, we predicted the presence of multiple PTM sites in BoCNGCs. Apart from evolutionarily conserved PTMs, other types of modification sites were detected in BoCNGCs, which diversified the functions and underlying mechanisms of CNGC-specific PTMs. Protein–protein interaction networks provide a base for systematic understanding of cellular processes that can be used to filter and assess the functional genomics data and provide an instinctive platform to annotate the

structures, functions and evolutionary properties of proteins [53]. Using two different approaches, and orthologous *Arabidopsis* CNGCs as a reference, we found that most CNGCs were associated with various protein–protein interaction networks involving CNGCs and other proteins related to light signalling [54], regulation of enzyme activities [55] and cellular processes [56], brassinosteroid signal transduction [57], and resistance against pathogens [58]. These analyses can offer new information for future experimental research and provide cross-species predictions for efficient interaction mapping [53]. Additionally, of the 26 BoCNGC genes, nine included target sites for diverse groups of novel and conserved miRNAs. These miRNA families are highly conserved in *Brassicaceae* species, where they are expressed in leaves, siliques, and flowers. These miRNAs are reported to function in regulation of genes related to growth (miR157/171/824) [59], *Brassica*-specific stomatal organization (miR824), pollen development (miR824) [60], abiotic stress tolerance, and plant–pathogen interactions (miR398) [61].

Gene duplications during evolution increase the genomic content and expand gene functions to optimise the adaptability of plants [25]. *Brassica oleracea* is an ancient polyploid, whose genome underwent a WGT event approximately 16 million years ago, after diverging from *A. thaliana*, followed by large-scale chromosomal rearrangements (i.e., re-diploidisation). As a member of the classical triangle of U [62], the assembled genome of *B. oleracea* (540 Mb) is larger than that of its sister species, *B. rapa* (312 Mb) [63] that diverged from a common ancestor nearly 4 million years ago [64]. The less number of CNGC genes in *Brassica* genomes suggest that most of the duplicated gene copies were lost post-polyploidization. Reversion of the few duplicated CNGC genes to single copy might be due to neutral loss of unnecessary duplicates over time. Another possible explanation could be that CNGC proteins participate in dosage sensitive interactions that is affected by the copy number of each protein subunit (gene balance hypothesis) [24]. Synteny analysis revealed that more than 80% of the BoCNGC genes are located in conserved syntenic blocks, which lost and gained some genes. These results are consistent with the findings of Liang et al. [65]. We presume that functionally redundant gene copies are reportedly lost after genome duplication events, while some copies of functionally important genes are kept [51]. Our findings suggest that the WGT and segmental duplication events were important for the expansion of the *B. oleracea* CNGC family, where tandem duplications only affected the expansion of Group IV-b. Altogether, the conservation of CNGC genes after substantial genome reshuffling suggests that these genes are crucial for plant development [66]. Finally, the detailed analyses of

gene expression in different tissues and under stress conditions further supported the importance of various *CNGC* genes for *B. oleracea* growth, development, and survival. To the best of our knowledge, this manuscript is the first to describe a comprehensive and systematic analysis of the *B. oleracea* *CNGC* gene family. The generated data may be useful for constructing protein–protein interaction networks and experimentally validating the miRNA targets, which regulate the development of *B. oleracea*. Besides, our results might help in understanding the functions of BoCNGCs related to the regulation of signal transduction pathways, and elucidate the expression profiles of the corresponding genes during plant development and stress responses. The results of the bioinformatics and comparative genomic analyses are also valuable for studying *CNGC* protein functions, with potential implications for the economic, agronomic, and ecological enhancement of *B. oleracea* and other *Brassica* species.

Conclusions

In conclusion, this work is the first comprehensive and systematic analyses of *CNGC* gene family in *B. oleracea*. There are 26 *CNGC* genes in *B. oleracea*, which are classified into 4 groups (I–IV) and fractionated into three sub-genomes; this gene family appears to have expanded through WGT, segmental and tandem duplication events; the BoCNGC gene family is under positive selection pressure. All the BoCNGC protein sequences contain a *CNGC* specific domain CNBD that comprises a PBC and a “hinge” region, featured by a stringent motif: LI]-X(2)-[GSE]-X-[VFIY]-X-G-X(0,1)-[DE]-L-L-X-W-X-[LQ]-X(10, 20)-S-X-[SAR]-X(7)-[VTI]-E-[AG]-F-X-L. This study provided comprehensive information about domain structure, exon-intron structure, and the phylogenetic tree and expression analysis of *CNGC* genes in Chinese cabbage. These data are useful to construct protein-protein interaction network and experimentally validate the miRNA targets, which regulates and induces multiple responses in *B. oleracea*. The bioinformatics analysis and comparative genomic analysis also provides valuable information in the study of *CNGC* protein functions for the improvement of the economic, agronomic, and ecological benefits of Chinese cabbage. Furthermore, this study assists to elucidate the functions of differentially expressed candidate genes in the regulation of signal transduction pathway, plant development and stress resistance in *B. oleracea*.

Methods

Identification of *Brassica oleracea* *CNGC* genes

To identify the *B. oleracea* *CNGC* genes, 20 *Arabidopsis* *CNGC* protein sequences obtained from TAIR10 (<https://www.arabidopsis.org/>) [67] were used as queries to perform a homology-based search of the Ensembl Plants

database (genome version v.2.1) [68]. This search was conducted with the default parameters of the BLASTP program. All non-redundant protein sequences were retrieved, and their domains were analysed with online servers: Simple Modular Architecture Research Tool (SMART) (<http://smart.embl-heidelberg.de/>) [69] and the Conserved Domains Database (CDD) (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [70]. The analyses were completed with the default cut-off parameters. Sequences containing the cNMP/CNBD (IPR000595) and transmembrane/ion transport protein (PF00520) domains as well as a plant *CNGC*-specific motif in the PBC and hinge region within the CNBD were recognized as *CNGC* proteins. The identified BoCNGC genes were named according to their positions in the phylogenetic tree.

Protein characterisation and amino acid properties

Details regarding gene and protein lengths as well as chromosomal locations were obtained from the Ensembl Plants database. Amino acid properties, including charge, molecular weight (kDa), aliphatic and instability indices, isoelectric points (pI), and grand average of hydropathy (GRAVY), were determined using the online available ProtParam tool (<http://web.expasy.org/protparam/>) [71]. The PTM sites were predicted with the ScanProsite web server (<http://prosite.expasy.org/scanprosite/>) [30].

Multiple sequence alignment and phylogenetic analysis

The identified *CNGC* proteins were aligned using the default settings of the ClustalX 2.0 program [72]. The conserved *CNGC*-specific domains were manually checked and shaded with the DNAMAN program (version 6.0.3.40; Lynnon Corporation, Quebec, Canada). The BoCNGC protein sequences were also aligned with *CNGC* sequences from *A. thaliana* and *B. rapa* (downloaded from the *Brassica* database; <http://brassicadb.org/brad/>) [73] using the default settings of the ClustalX 2.0 program. The alignments were viewed with the GeneDoc program [74]. A phylogenetic tree was constructed using the maximum likelihood method of MEGA 6.0 (1000 bootstrap replications) [75].

Chromosomal locations and gene duplication events

Details regarding the chromosomal locations of the BoCNGC genes were obtained from the Ensembl Plants database. The Plant Genome Duplication Database [76] was searched to identify segmentally duplicated genes. BoCNGC genes were defined as tandemly duplicated if the distance between the homologous loci was <50 kb [65]. The syntenic relationships among BoCNGCs, AtCNGCs, and BrCNGCs were evaluated using the Search Syntenic Genes tool in Bolbase [77].

Gene structure, motif composition, and prediction of three-dimensional models

Gene exon/intron structures were predicted with the Gene Structure Display Server (version 2.0) [78], with genomic and coding sequences as the input data. The conserved motifs in the CNGC sequences were identified using the Multiple Expectation Maximization for Motif Elicitation suite and the Motif Alignment and Search Tool [28] with the following parameters: optimal motif width: 6–200; maximum number of different motifs: 10. The detected motifs were annotated with Pfam [29]. Gene ontology enrichment analysis was performed using Blast2GO (v.3.3.5) [79].

Analysis of microRNA target sites and protein–protein interactions

The *B. oleracea* miRNA sequences obtained from the miRBase database at <http://mirbase.org/> [80]. To detect potential miRNA target sites within the *BoCNGC* genes, the obtained miRNAs were analysed with the psRNATarget server (<http://plantgrn.noble.org/psRNA-Target/>) [44]. The information about protein-protein interaction, and available mutant information for *Arabidopsis* CNGC-encoded proteins was obtained from STRING (v10) [36] and AtPID (<http://www.megabionet.org/atpid/webfile/query.php>).

Analysis of *BoCNGC* transcriptome data

To investigate the *BoCNGC* expression profiles, we used the Illumina RNA-sequencing data available in the Gene Expression Omnibus database (accession number GSE42891) [24]. Transcript abundance was calculated as fragments per kilobase of exon model per million mapped reads, and the resulting values were \log_2 transformed. A hierarchical cluster was created and a heat map was generated with R language program [81].

Experimental conditions and quantitative real-time polymerase chain reaction assay

We used a quantitative real-time polymerase chain reaction (qRT-PCR) to quantify the *BoCNGC* expression levels in response to biotic (bacterial pathogen) and abiotic (cold) stresses. Cabbage (*B. oleracea* var. *capitata* L.) seedlings were grown for 25 days in a greenhouse at 23 ± 2 °C under natural light. For the cold stress treatment, seedlings were incubated at 4 °C for 24 h. For the bacterial infection, Xcc was first cultured in medium B [82] at 28 °C. Cells were collected by centrifugation, re-suspended in sterilized distilled water, and adjusted to an optical density at 600 nm of 0.1. The midvein of the first fully opened leaf (just above the petiole) was inoculated with the Xcc suspension using a 1-ml syringe. Sterilized ddH₂O was used as the control solution. The treated plants were returned to the greenhouse and sampled 24 h later. The extraction

of RNA and synthesis of cDNA were completed as previously described [20]. Gene-specific primers were designed with Primer 5.0 (Additional file 23). The qRT-PCR was conducted using a StepOne Real-Time PCR System (Applied Biosystems, USA) and SYBR Premix Ex Taq reagents (TAKARA, Japan) as described by Kabouw et al. [83]. Finally, the $2^{-\Delta\Delta C_t}$ method [84] was used to calculate the relative gene expression values, which were subsequently transformed to \log_2 - expression ratios and plotted in figures. Each experiment was performed with three technical replicates. The *Actin* gene (AF044573) was used as an endogenous control.

Statistical analysis

The RT-qPCR expression data was subjected to analysis of variance (ANOVA) using computer statistical package (SAS software SAS Institute, Cary, NC). Least significant difference (LSD) test at $p \leq 0.01$ was used to check the significant differences between the expression levels of different *BoCNGC* genes compared to control.

Additional files

Additional file 1: List of truncated gene accessions discarded during preliminary investigation. (XLSX 9 kb)

Additional file 2: Multiple sequence alignment of CNGC proteins from *B. oleracea*, *B. rapa* and *A. thaliana*. (PDF 2222 kb)

Additional file 3: Phylogenetic tree of CNGC proteins from *B. oleracea* (encoded by *BoCNGCs*). A multiple sequence alignment was performed using ClustalX 2.0 program with default settings. Maximum likelihood (ML) tree was created with MEGA 6.0, under the Jones-Taylor-Thornton (JTT) model. The bootstrap values from 1000 resampling are given at each node. (PDF 201 kb)

Additional file 4: Phylogenetic tree of CNGC genes from *Arabidopsis* (*AtCNGCs*). A multiple sequence alignment was performed using ClustalX 2.0 program with default settings. Maximum likelihood (ML) tree was created with MEGA 6.0, under the Jones-Taylor-Thornton (JTT) model. The bootstrap values from 1000 resampling are given at each node. (PDF 180 kb)

Additional file 5: Syntenic ancestral block structure between *A. thaliana* and three sub-genomes of *B. oleracea* and *B. rapa*. (XLSX 10 kb)

Additional file 6: Synteny of *BoCNGC* in other plant species. (XLSX 18 kb)

Additional file 7: Primary domain architecture of *BoCNGC* proteins. Information about domain annotation is obtained from SMART database. (PDF 339 kb)

Additional file 8: Multiple sequence alignment of *BoCNGC* proteins. Multiple sequence alignment was performed by clustalX2 and viewed by GeneDoc software package. (PDF 767 kb)

Additional file 9: Multiple sequence alignment of CNGC-encoded proteins of *Arabidopsis* and *B. oleracea*. Multiple sequence alignment was performed by clustal X2 and viewed by GeneDoc software package. (PDF 1208 kb)

Additional file 10: Multiple sequence alignment of CNGC-encoded proteins of *B. oleracea* and *B. rapa*. Multiple sequence alignment was performed by clustal X2 and viewed by GeneDoc software package. (PDF 1436 kb)

Additional file 11: Schematic diagram showing the structures of *Arabidopsis* CNGC family genes. The exons-introns indicated as red boxes and black lines respectively, and the intron phases are displayed as numbers [0, 1 and 2]. The lengths of each exon and intron can be mapped to the scale given in the bottom. (PDF 154 kb)

Additional file 12: Functional annotation of the identified conserved MEME motifs. (XLSX 10 kb)

Additional file 13: Web logos of MEME-identified conserved functional motifs in BoCNGC proteins. The heights of the amino acids indicates the degree of conservation. (PDF 423 kb)

Additional file 14: KO pathway associated with plant-pathogen interaction (K05391). The pathway map was obtained from <http://www.kegg.jp/kegg/kegg1.html>. Details of BoCNGC genes allocated to K05391 pathway are given in Additional file 20. (PDF 126 kb)

Additional file 15: Table showing the details of protein-protein interaction, and available mutant information for *Arabidopsis* CNGC-encoded proteins. The information was obtained from AtPID. (XLSX 37 kb)

Additional file 16: The potential miRNA targets in the set of 26 BoCNGC transcripts using cut-off threshold of 5 in the search parameters. (XLSX 14 kb)

Additional file 17: GO term enrichment analysis of BoCNGC genes for Molecular function (MF), Biological process (BP) and Cellular component (CC). (XLSX 17 kb)

Additional file 18: Distribution of BoCNGC genes in major functional terms (GO terms Level 2) for categories Molecular Function (a), Biological Process (b) and Cellular Component (c). The details are given in Additional file 19. (PDF 97 kb)

Additional file 19: GO term enrichment analysis at level 2 for category: P: Biological process, F: Molecular function and C: Cellular component. (XLSX 9 kb)

Additional file 20: Reference KO pathway associated with BoCNGC genes. The pathway map was obtained from <http://www.kegg.jp/kegg/kegg1.html>. (XLSX 10 kb)

Additional file 21: Cumulative values of expression for *Arabidopsis* CNGC genes in different developmental samples. The expression data for 21 days old of wild type and mutant plants was obtained from Schmid et al. [48]. The information about different genotype mutants is given below the figures. (PDF 225 kb)

Additional file 22: Cumulative values of expression for *Arabidopsis* CNGC genes in response to pathogen (biotic) and cold (Abiotic) stress. The expression data for 21 days old of wild type and mutant plants was obtained from Schmid et al. [48]. (PDF 294 kb)

Additional file 23: List of primers used for gene expression via qRT-PCR. (XLSX 12 kb)

Abbreviations

AMD: Amidation; ASN: N-glycosylation; CaM: calmodulin; CaMBD: CaM-binding domain; cAMP: cyclic adenosine monophosphate; cGMP: cyclic guanosine monophosphate; CK2: Casein kinase II; CNBD: cNMP-binding domain; CNGC: Cyclic nucleotide-gated ion channel; IQ: Isoleucine-glutamine; IT: Ion transport; LEU: Leucine zipper pattern; MYR: N-myristoylation; PBC: Phosphate-binding cassette; pI: Isoelectric point; PKC: Protein kinase C; PTM: Post-translational modification; RGD: Cell attachment sequence; TYR: Tyrosine kinase; WGD: Whole-genome duplication; Xcc: *Xanthomonas campestris* pv. *campestris*

Acknowledgements

We thank Prof. Qing-yao Shu for his critical inputs and assistance during this study.

Funding

This research was financially supported by the Ministry of Science and Technology, China (grant No.: SQ2015IM3600010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The sequence datasets analysed during the current study are publicly available in the Ensembl Genomes [http://plants.ensembl.org/Brassica_oleracea/Info/Index]. The transcriptomic data of BoCNGC genes used in current analyses are available in the Gene Expression Omnibus database (accession number GSE42891).

Authors' contributions

KUK and ZN designed the study and conceptualised the methodology. KK and EA collected the data and completed the bioinformatics analyses along with KUK and ZN. RU and AA conducted the qRT-PCR experiments with assistance from KK. KUK and ZN analysed the data and wrote the manuscript with critical inputs from X-LR and Q-YS. X-LR and Q-YS revised the whole manuscript according to suggestions by referees, and supervised this study. All authors reviewed the manuscript at every stage. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The Cabbage seeds were provided by Zhejiang Key Laboratory of Crop Gene Resources, College of Agriculture and Biotechnology, Zhejiang University, China, and no permissions are needed to obtain the material. Our study fully complies with institutional regulations.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹State Key Laboratory of Rice Biology, Institute of Crop Science, Zhejiang University, Hangzhou 310058, China. ²Molecular Genetics Key Laboratory of China Tobacco, Guizhou Academy of Tobacco Science, Guiyang 550081, China. ³Wuxi Hupper Bioseed Technology Academy Ltd., Wuxi 214000, China. ⁴Department of Biotechnology, BUITEMS, Quetta, Pakistan. ⁵Department of Biological sciences, College of Education and Science, Albaydaa University, Rada'a, Yemen. ⁶Department of Environmental Sciences, Quaid -i- Azam University, Islamabad, Pakistan. ⁷Institute of Crop Sciences, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310029, China. ⁸Guizhou Academy of Tobacco Science, Longtanba Road No. 29, Guanshanhu District, Guiyang (550081), Guizhou, People's Republic of China.

Received: 12 July 2017 Accepted: 31 October 2017

Published online: 13 November 2017

References

1. Wu M, Li Y, Chen D, Liu H, Zhu D, Xiang Y. Genome-wide identification and expression analysis of the IQD gene family in moso bamboo (*Phyllostachys edulis*). *Sci Rep*. 2016;6:24520.
2. Takáč T, Vadovič P, Pečan T, Luptovčíak I, Šamajová O, Šamaj J. Comparative proteomic study of *Arabidopsis* mutants mpk4 and mpk6. *Sci Rep*. 2016;6:28306.
3. DeFalco TA, Marshall CB, Munro K, Kang H-G, Moeder W, Ikura M, Snedden WA, Yoshioka K. Multiple Calmodulin-binding sites positively and negatively regulate *Arabidopsis* Cyclic Nucleotide-gated Channel12. *Plant Cell*. 2016;28(7):1738–51.
4. Saand MA, Xu Y-P, Munyampundu J-P, Li W, Zhang X-R, Cai X-Z. Phylogeny and evolution of plant cyclic nucleotide-gated ion channel (CNGC) gene family and functional analyses of tomato CNGCs. *DNA Res*. 2015;22(6):471–83.
5. Mäser P, Thomine S, Schroeder JI, Ward JM, Hirschi K, Sze H, Talke IN, Amtmann A, Maathuis FJM, Sanders D. Phylogenetic relationships within cation transporter families of *Arabidopsis*. *Plant Physiol*. 2001;126(4):1646–67.
6. Borsics T, Webb D, Andeme-Ondzighi C, Staehelin LA, Christopher DA. The cyclic nucleotide-gated calmodulin-binding channel AtCNGC10 localizes to the plasma membrane and influences numerous growth responses and starch accumulation in *Arabidopsis thaliana*. *Planta*. 2007;225(3):563–73.
7. Christopher DA, Borsics T, Yuen CY, Ullmer W, Andeme-Ondzighi C, Andres MA, Kang B-H, Staehelin LA. The cyclic nucleotide gated cation channel AtCNGC10 traffics from the ER via Golgi vesicles to the plasma membrane of *Arabidopsis* root and leaf cells. *BMC Plant Biol*. 2007;7(1):48.
8. Yuen CC, Christopher DA. The group IV-A cyclic nucleotide-gated channels, CNGC19 and CNGC20, localize to the vacuole membrane in *Arabidopsis thaliana*. *AoB Plants*. 2013;5:plt012.

9. Charpentier M, Sun J, Martins TV, Radhakrishnan GV, Findlay K, Soumpourou E, Thouin J, Véry A-A, Sanders D, Morris RJ. Nuclear-localized cyclic nucleotide-gated channels mediate symbiotic calcium oscillations. *Science*. 2016;352(6289):1102–5.
10. Newton RP, Smith CJ. Cyclic nucleotides. *Phytochemistry*. 2004;65(17):2423–37.
11. Kaplan B, Sherman T, Fromm H. Cyclic nucleotide-gated channels in plants. *FEBS Lett*. 2007;581(12):2237–46.
12. Ma W, Berkowitz GA. Cyclic nucleotide gated channel and Ca²⁺-mediated signal transduction during plant senescence signaling. *Plant Signal Behav*. 2011;6(3):413–5.
13. Ma W, Smigel A, Walker RK, Moeder W, Yoshioka K, Berkowitz GA. Leaf senescence signaling: the Ca²⁺-conducting Arabidopsis cyclic nucleotide gated channel2 acts through nitric oxide to repress senescence programming. *Plant Physiol*. 2010;154(2):733–43.
14. Zelman AK, Dawe A, Gehring C, Berkowitz GA. Evolutionary and structural perspectives of plant cyclic nucleotide-gated cation channels. *Front Plant Sci*. 2012;3(195):95.
15. Guo KM, Babourina O, Christopher DA, Borsics T, Rengel Z. The cyclic nucleotide-gated channel, AtCNGC10, influences salt tolerance in Arabidopsis. *Physiol Plant*. 2008;134(3):499–507.
16. Nawaz Z, Kakar KU, Saand MA, Shu Q-Y. Cyclic nucleotide-gated ion channel gene family in rice, identification, characterization and experimental analysis of expression response to plant hormones, biotic and abiotic stresses. *BMC Genomics*. 2014;15(1):853.
17. Saand MA, Xu Y-P, Li W, Wang J-P, Cai X-Z. Cyclic nucleotide gated channel gene family in tomato: genome-wide identification and functional analyses in disease resistance. *Front Plant Sci*. 2015;6:303.
18. Chen J, Yin H, Gu J, Li L, Liu Z, Jiang X, Zhou H, Wei S, Zhang S, Wu J. Genomic characterization, phylogenetic comparison and differential expression of the cyclic nucleotide-gated channels gene family in pear (*Pyrus bretschneideri* Rehd.). *Genomics*. 2015;105(1):39–52.
19. Zelman AK, Dawe A, Berkowitz GA. Identification of cyclic nucleotide gated channels using regular expressions. In: Gehring C, editor. *Cyclic nucleotide signaling in plants: methods and protocols*. Totowa: Humana Press; 2013. p. 207–24.
20. Almoneafy AA, Kakar KU, Nawaz Z, Li B, Chun-lan Y, Xie G-L. Tomato plant growth promotion and antibacterial related-mechanisms of four rhizobacterial bacillus strains against *Ralstonia solanacearum*. *Symbiosis*. 2014;63(2):59–70.
21. Chin K, DeFalco TA, Moeder W, Yoshioka K. The Arabidopsis cyclic nucleotide-gated ion channels AtCNGC2 and AtCNGC4 work in the same signaling pathway to regulate pathogen defense and floral transition. *Plant Physiol*. 2013;163(2):611–24.
22. Finka A, Cuendet AFH, Maathuis FJ, Saidi Y, Goloubinoff P. Plasma membrane cyclic nucleotide gated calcium channels control land plant thermal sensing and acquired thermotolerance. *Plant Cell*. 2012;24(8):3333–48.
23. Warwick SJ, Francis A, Al-Shehbaz IA. Brassicaceae: species checklist and database on CD-Rom. *Plant Syst Evol*. 2006;259(2–4):249–58.
24. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, et al. The Brassica Oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun*. 2014;5:3930.
25. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon–intron structure. *Proc Natl Acad Sci*. 2012;109(4):1187–92.
26. Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica Oleracea. *Genome Bio*. 2014;15(6):R77.
27. Lysak MA, Koch MA, Pecinka A, Schubert I. Chromosome triplication found across the tribe Brassicaceae. *Genome Res*. 2005;15(4):516–25.
28. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37(suppl_2):W202–8.
29. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42(D1):D222–30.
30. De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*. 2006;34(suppl_2):W362–5.
31. Leppänen T, Tuominen RK, Moilanen E. Protein Kinase C and its inhibitors in the regulation of inflammation: inducible nitric oxide Synthase as an example. *Basic Clin Pharmacol Toxicol*. 2014;114(1):37–43.
32. Mulekar JJ, Bu Q, Chen F, Huq E. Casein kinase II α subunits affect multiple developmental and stress-responsive pathways in Arabidopsis. *Plant J*. 2012; 69(2):343–54.
33. Traverso JA, Meinel T, Giglione C. Expanded impact of protein N-myristoylation in plants. *Plant Signal Behav*. 2008;3(7):501–2.
34. Strasser R. Plant protein glycosylation. *Glycobiology*. 2016;26(9):926–39.
35. Lai S, Pelech S. Regulatory roles of conserved phosphorylation sites in the activation T-loop of the MAP kinase ERK1. *Mol Biol Cell*. 2016;27(6):1040–50.
36. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2014;43(D1):D447–52.
37. Schläfli P, Tröger J, Eckhardt K, Bortler E, Spielmann P, Wenger RH. Substrate preference and phosphatidylinositol monophosphate inhibition of the catalytic domain of the per-Arnt-Sim domain kinase PASKIN. *FEBS J*. 2011; 278(10):1757–68.
38. Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. *Nature*. 2010;466(7302):68–76.
39. Chen J, Lalonde S, Obrdlik P, Noorani Vatani A, Parsa SA, Vilarino C, Revuelta JL, Frommer WB, Rhee SY. Uncovering Arabidopsis membrane protein Interactome enriched in transporters using mating-based split Ubiquitin assays and classification models. *Front Plant Sci*. 2012;3:124.
40. Frietsch S, Wang Y-F, Sladek C, Poulsen LR, Romanowsky SM, Schroeder JI, Harper JF. A cyclic nucleotide-gated channel is essential for polarized tip growth of pollen. *Proc Natl Acad Sci U S A*. 2007;104(36):14531–6.
41. Ladwig F, Dahlke RI, Stührowohldt N, Hartmann J, Harter K, Sauter M. Phytosulfokine regulates growth in Arabidopsis through a response module at the plasma membrane that includes CYCLIC NUCLEOTIDE-GATED CHANNEL17, H⁺-ATPase, and BAK1. *Plant Cell*. 2015;27(6):1718–29.
42. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41(Database issue):D808–15.
43. Witkos TM, Koscińska E, Krzyżosiak WJ. Practical aspects of microRNA target prediction. *Curr Mol Med*. 2011;11(2):93–109.
44. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res*. 2011;39(suppl_2):W155–9.
45. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2015;44(D1):D457–62.
46. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. A gene expression map of Arabidopsis Thaliana development. *Nat Genet*. 2005;37(5):501–6.
47. Ma W, Ali R, Berkowitz GA. Characterization of plant phenotypes associated with loss-of-function of AtCNGC1, a plant cyclic nucleotide gated cation channel. *Plant Physiol Biochem*. 2006;44(7):494–505.
48. Khaldi N, Shields DC. Shift in the isoelectric-point of milk proteins as a consequence of adaptive divergence between the milks of mammalian species. *Biol Direct*. 2011;6(1):40.
49. Schwartz R, Ting CS, King J. Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res*. 2001;11(5):703–9.
50. Gitlin I, Carbeck JD, Whitesides GM. Why are proteins charged? Networks of charge–charge interactions in proteins measured by charge ladders and capillary electrophoresis. *Angew Chem Int Ed*. 2006;45(19):3022–60.
51. Duan G, Walther D. The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol*. 2015;11(2):e1004049.
52. Webster DE, Thomas MC. Post-translational modification of plant-made foreign proteins; glycosylation and beyond. *Biotechnol Adv*. 2012;30(2):410–8.
53. Schwartz AS, Yu J, Gardenour KR, Finley RL, Ideker T. Cost effective strategies for completing the Interactome. *Nat Methods*. 2009;6(1):55–61.
54. Bauer D, Viczián A, Kircher S, Nobis T, Nitschke R, Kunkel T, Panigrahi KCS, Ádám É, Fejes E, Schäfer E. Constitutive photomorphogenesis 1 and multiple photoreceptors control degradation of phytochrome interacting factor 3, a transcription factor required for light signaling in Arabidopsis. *Plant Cell*. 2004;16(6):1433–45.
55. Cohen P. *Control of enzyme activity, illustrated edn*. Berlin: Springer Science & Business Media; 2013.
56. Banerjee J, Magnani R, Nair M, Dirk LM, DeBolt S, Maiti IB, Houtz RL. Calmodulin-mediated signal transduction pathways in Arabidopsis are fine-tuned by methylation. *Plant Cell*. 2013;25(11):4493–511.

57. Sun Y, Li L, Macho AP, Han Z, Hu Z, Zipfel C, Zhou J-M, Chai J. Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science*. 2013;342(6158):624–8.
58. Murata Y, Mori IC, Munemasa S. Diverse stomatal signaling and the signal integration mechanism. *Annu Rev Plant Biol*. 2015;66:369–92.
59. Lukasik A, Pietrykowska H, Paczek L, Szwejkowska-Kulinska Z, Zielenkiewicz P. High-throughput sequencing identification of novel and conserved miRNAs in the Brassica Oleracea leaves. *BMC Genomics*. 2013;14(1):801.
60. Song JH, Yang J, Pan F, Jin B. Differential expression of microRNAs may regulate pollen development in Brassica Oleracea. *Gen Mol Res*. 2015;14(4):15024–34.
61. He X-F, Fang Y-Y, Feng L, Guo H-S. Characterization of conserved and novel microRNAs and their targets, including a TuMV-induced TIR–NBS–LRR class R gene-derived novel miRNA in Brassica. *FEBS Lett*. 2008;582(16):2445–52.
62. Nagaharu U. Genome analysis in Brassica with special reference to the experimental formation of *B. Napus* and peculiar mode of fertilization. *Jpn J Bot*. 1935;7:389–452.
63. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B. Early allopolyploid evolution in the post-Neolithic Brassica Napus oilseed genome. *Science*. 2014;345(6199):950–3.
64. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S. The Brassica Oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun*. 2014;5:3930.
65. Liang Y, Xiong Z, Zheng J, Xu D, Zhu Z, Xiang J, Gan J, Raboanatahiry N, Yin Y, Li M. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in Brassica Napus. *Sci Rep*. 2016;6:24265.
66. Cheng F, Mandáková T, Wu J, Xie Q, Lysak MA, Wang X. Deciphering the diploid ancestral genome of the mesohexaploid Brassica Rapa. *Plant Cell*. 2013;25(5):1541–54.
67. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M. The Arabidopsis information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*. 2012;40(D1):D1202–10.
68. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C. Ensembl genomes 2016: more genomes, more complexity. *Nucleic Acids Res*. 2016;44(D1):D574–80.
69. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*. 2015;43(D1):D257–60.
70. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 2014;43(D1):D222–6.
71. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. Totowa: Humana Press; 2005.
72. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–8.
73. Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biol*. 2011; 11(1):1.
74. Nicholas KB, Nicholas HBJ: GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author; 1997.
75. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725–9.
76. Lee T-H, Tang H, Wang X, Paterson AH. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res*. 2013;41(D1):D1152–8.
77. Yu J, Zhao M, Wang X, Tong C, Huang S, Tehrim S, Liu Y, Hua W, Liu S. Bolbase: a comprehensive genomics database for Brassica Oleracea. *BMC Genomics*. 2013;14(1):1.
78. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31(8):1296–7.
79. Conesa A, Götz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*. 2008;2008:12.
80. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014;42(D1): D68–73.
81. RCoreTeam. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014.
82. King EO, Ward MK, Raney DE. Two simple media for the demonstration of pyocyanin and fluorescein. *J Lab Clin Med*. 1954;44(2):301–7.
83. Kabouw P, Biere A, van der Putten WH, van Dam NM. Intra-specific differences in root and shoot Glucosinolate profiles among white cabbage (*Brassica Oleracea* Var. *Capitata*) cultivars. *J Agric Food Chem*. 2010;58(1):411–7.
84. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ΔΔCT method. *Methods*. 2001;25(4): 402–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

